Projekt realizowany
z Narodowym Bankiem Polskim
w ramach programu edukacji ekonomicznej

NBP Narodowy Bank Polski

SIGMA KWADRAT

LUBELSKI KONKURS STATYSTYCZNO-DEMOGRAFICZNY

Projekt realizowany
z Narodowym Bankiem Polskim
w ramach programu edukacji ekonomicznej

Pojęcie i metody badań statystycznych

Projekt realizowany
z Narodowym Bankiem Polskim
w ramach programu edukacji ekonomicznej

Bibliografia

Materiały przygotowano na podstawie podręczników:

Mieczysław Sobczyk, *Statystyka opisowa*, Wydawnictwo C.H. Beck, 2010

Mieczysław Sobczyk, *Statystyka. Aspekty praktyczne i teoretyczne*, Wydawnictwo UMCS w Lublinie, Lublin 2006

oraz:

Jerzy Zdzisław Holze, *Demografia*, PWE, Warszawa, 2003

Geneza statystyki

Termin statystyka wywodzi się od łacińskiego słowa **status** czyli państwo.

W piśmiennictwie słowo **statystyka** po raz pierwszy zostało użyte w połowie XVIII wieku przez Gottfrieda Achenwalla (1719 – 1772), profesora uniwersytetu w Marburgu, a następnie w Getyndze. Do połowy XIX w., określenie **statystyka** oznaczało zbiór danych liczbowych dotyczących stanu państwa, a następnie wszelkie gromadzone informacje opisujące świat zewnętrzny. Dynamiczny rozwój statystyki jako narzędzia wykorzystywanego do opisu zjawisk gospodarczych (tablice bilansowe, warunki bytu ludności itp.) datuje się od czasów A. Quetleta (1796 – 1874). W okresie tym statystykę zaczęto traktować jako naukę, czego wyrazem było powołanie w 1834 r. w Anglii Królewskiego Towarzystwa Statystycznego oraz zwołanie w Brukseli I Międzynarodowego Kongresu Statystycznego (1854 r.).

Pojęcie i metody badań statystycznych

Współcześnie termin **statystyka** ma wiele znaczeń. W mowie potocznej używa się go w różnym znaczeniu, np.. jako nazwy:

1. Zbioru danych liczbowych, przedstawiających kształtowanie się określonych zjawisk i procesów. W tym znaczeniu mówimy np. o statystyce ludności, transportu, rolnictwa itp.
2. Wszelkich prac związanych z gromadzeniem i opracowywaniem danych liczbowych.
3. Pewnych charakterystyk opisowych obliczanych ze zbiorowości próbnych. W tym sensie statystyką jest np. średnia arytmetyczna z próby, odchylenie standardowe z próby, współczynnik korelacji liniowej Pearsona z próby.
4. Dyscypliny naukowej mającej własne, specyficzne metody badawcze. W tym znaczeniu statystykę definiuje się jako naukę o „ilościowych metodach badania prawidłowości występujących w zjawiskach masowych scharakteryzowanych za pomocą liczb”.

Pojęcie i metody badań statystycznych

STATYSTYKĘ (ogólną) jako dyscyplinę naukową możemy podzielić na:

- ***statystykę opisową*** zajmującą się programowaniem badań statystycznych oraz metod obserwacji, a także prezentacją materiału statystycznego,

- ***statystykę matematyczną*** zajmującą się metodami wnioskowania o całej zbiorowości generalnej na podstawie badań.

System informacyjny statystyki publicznej

Zasady gromadzenia danych i prowadzenia badań statystycznych reguluje Ustawa o statystyce publicznej z 29 czerwca 1995 r. (Dz. U., Nr 88, poz. 439). Całokształt działań związanych ze zbieraniem informacji statystycznych, ich gromadzeniem, przechowywaniem, opracowywaniem, ogłaszaniem, udostępnianiem i rozpowszechnianiem wyników badań statystycznych jako oficjalnych danych tworzy **System Informacyjny Statystyki Publicznej (SISP)**.

Zadania wynikające z funkcjonowania SISP realizowane są przez służby statystyki publicznej, tzn. Główny Urząd Statystyczny (GUS) oraz podległe mu urzędy statystyczne. GUS, utworzony w 1918 r., jest centralnym organem administracji państwowej, którego zadaniem są m.in.: prowadzenie i udostępnianie wyników badań statystycznych, międzynarodowa współpraca

w zakresie statystyki, rozwijanie metodologii badań statystycznych oraz ustalanie i aktualizacja podstawowych definicji, kodów, klasyfikacji i nomenklatur (tj. zbioru nazw i terminów używanych w statystyce).

Zadania statystyki publicznej

Do zadań służb statystyki publicznej należy także prowadzenie,w skali krajowej, dwóch rejestrów: urzędowego (zwanego też administracyjnym) oraz podmiotów gospodarki narodowej (REGON).

Rejestr urzędowy jest wykazem podmiotów informującym o ich działalności. Jest prowadzony przez sądy i organy administracji publicznej na podstawie aktów wykonawczych do ustaw (np. rejestry podatników, działalności gospodarczej, udzielonych zezwoleń i koncesji, gruntów, budynków).

REGON to powszechny system identyfikacji i kodowania jednostek gospodarki narodowej. W rejestrze tym ujmowane są dane o wszystkich osobach prawnych, jednostkach nieposiadających osobowości prawnej, indywidualnych gospodarstwach rolnych oraz osobach fizycznych prowadzących działalność gospodarczą.

Dowodem wpisu do REGON-u jest nadanie określonemu podmiotowi numeru identyfikacyjnego, zawierającego zakodowane informacje o jego cechach.

Pojęcie i metody badań statystycznych

Celem analiz statystycznych jest wykrycie prawidłowości rządzących badanymi zjawiskami masowymi. Obserwowane w zjawiskach masowych prawidłowości są efektem oddziaływania dwóch grup przyczyn: głównych (systematycznych) i ubocznych (przypadkowych). Przyczyny główne oddziałują na badane zjawisko w sposób jednokierunkowy, kształtując tzw. składnik systematyczny zjawiska masowego. Dzięki nim ujawnia się charakterystyczna dla danego zjawiska masowego prawidłowość.

Pojęcie i metody badań statystycznych

Przyczyny uboczne występują w dużej ilości i działają różnokierunkowo. W obserwowanej zbiorowości wywołują one odchylenia od prawidłowości, które ujawniają się w indywidualnych przypadkach. Rezultatem oddziaływania przyczyn ubocznych jest przypadkowy składnik zjawiska masowego. Skutki działania przyczyn głównych uwidaczniają się tym wyraźniej, im liczniejsza jest poddana badaniu zbiorowość. Mówimy wówczas o działaniu **prawa wielkich liczb**. Treść tego prawa można ująć następująco: przy dostatecznie dużej liczbie obserwacji dotyczących określonego zjawiska skutki przyczyn ubocznych redukują się, natomiast ujawniają się prawidłowości.

Sama masowość nie decyduje o możliwości stosowania w badaniach metod statystycznych. Metody te mogą być wykorzystywane jedynie do badania takich zjawisk masowych, które dotyczą zbiorowości składających się z elementów o podobnych – ale nie identycznych – właściwościach.

Pojęcie i metody badań statystycznych

Za podstawowe zadanie statystyki uznaje się dostarczanie wiarygodnych informacji niezbędnych do podejmowania różnego rodzaju decyzji. Jeśli decyzje te mają być skuteczne i efektywne, to powinny bazować na bogatym, poprawnie zgromadzonym i wszechstronnie przeanalizowanym zbiorze informacji.

Podstawowe **funkcje** spełniane przez statystykę:

- ✓ informacyjna – dająca pełny i obiektywny obraz badanych zjawisk,
- ✓ analityczna – dzięki której możliwe jest określenie czynników kształtujących konkretne procesy i zjawiska,
- ✓ prognostyczna – pozwalająca na przewidywanie kierunku rozwoju analizowanych zjawisk.

Pojęcie i metody badań statystycznych

Badanie statystyczne prowadzone jest zawsze na pewnej zbiorowości (osób, rzeczy, zjawisk, faktów itp.). Zbiór dowolnych elementów podobnych – ale nie identycznych – pod względem określonych właściwości i poddanych badaniu statystycznemu nazywamy **zbiorowością (populacją generalną)**. Elementy składowe zbiorowości poddane bezpośredniej obserwacji lub pomiarowi noszą nazwę jednostek statystycznych, jednostek badania lub jednostek obserwacji. Zbiorowości powinny być ściśle określone pod względem **rzeczowym** (kto lub co jest przedmiotem badania), **przestrzennym** (gdzie znajduje się zbiorowość) oraz **czasowym** (jakiego momentu lub okresu dotyczy badanie).

Pojęcie i metody badań statystycznych

Zbiorowości można wyodrębniać według różnych kryteriów. I tak, ze względu na kryterium czasu wyróżnia się zbiorowości **statyczne** (np. ludność Polski według stanu na dzień 31.12. ubiegłego roku) i **dynamiczne** (np. samochody zarejestrowane w Polsce w ubiegłym roku). Zbiorowości **statyczne** tworzą zatem jednostki, które istniały, istnieją lub będą istniały w ściśle określonym momencie. Zbiorowości **dynamiczne** składają się z jednostek obserwowanych w pewnym przedziale czasu. Według kryterium liczebności, wyróżniamy zbiorowości **skończenie** i **nieskończenie liczne**. Pierwsze z nich składają się z przeliczalnej liczby jednostek (np. studenci UMCS według stanu na 31.12. danego roku), drugie zaś – tworzą elementy o nieprzeliczalnej liczbie (np. zbiorowość organizmów żywych na kuli ziemskiej). Wyróżnia się również zbiorowości **jednorodne (**np. wszystkie dzieci płci męskiej urodzone w danym dniu) i **niejednorodne ((**np. wszystkie urodzone w danym dniu dzieci).

Klasyfikacje zbiorowości statystycznych

zbiorowość statystyczna

- **skończona** – ma skończoną liczbę jednostek, np. 70 pracowników firmy
- **nieskończona** – ma nieskończoną lub niemożliwą do ustalenia liczbę jednostek statystycznych, np. zbiorowość mikroorganizmów albo klienci odwiedzający centrum handlowe
- **jednowymiarowa** – badana ze względu na jedną cechę, np. pracownicy badani ze względu na miesięczną płacę
- **wielowymiarowa** – badana jednocześnie ze względu na kilka cech, np. pracownicy firmy, w przypadku badania zależności miesięcznej płacy od stażu pracy
- **względnie jednorodna** – jej podzbiorowości mało różnią się własnościami, np. zbiorowość gospodarstw dwuosobowych badana ze względu na tygodniowe wydatki na żywność
- **niejednorodna** – jej podzbiorowości wyraźnie różnią się własnościami, np. zbiorowość gospodarstw domowych o różnej wielkości, badana ze względu na tygodniowe wydatki na żywność
- **statyczna** – wszystkie jednostki statystyczne pochodzą z tego samego okresu, np. firmy farmaceutyczne zarejestrowane we wrześniu 2003 r., spółki na giełdzie notowane 14.06.2006 r.
- **dynamiczna** – jednostki statystyczne pochodzą z różnych okresów, np. kolejne sesje giełdowe w czerwcu 2006 r.

Źródło: A. Maksimowicz – Ajchel, *Wstęp do statystyki. Metody opisu statystycznego*, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 2007, s. 15

Pojęcie i metody badań statystycznych

Jednostki statystyczne tworzące określoną populacje generalną charakteryzują się pewnymi własnościami. Własności te nazywamy **cechami statystycznymi**. Cechy statystyczne można podzielić na **stałe** (wspólne dla wszystkich jednostek badanej zbiorowości) oraz **zmienne**, dzięki którym poszczególne jednostki różnią się między sobą. Cechy stałe (rzeczowe, przestrzenne i czasowe) nie podlegają badaniu statystycznemu, a jedynie służą do precyzyjnego zdefiniowania zbiorowości statystycznej. Analizie statystycznej poddawane są tylko cechy zmienne.

Cechy, których warianty podawane są w sposób opisowy nazywamy cechami **niemierzalnymi (jakościowymi)**. Cechy, których warianty są wyrażane za pomocą liczb, określa się mianem cech **mierzalnych (ilościowych)**.

Cechy jakościowe, których natężenie określone jest przez przymiotniki (np. ocena wiadomości ucznia: bardzo dobra, dobra, dostateczna, niedostateczna), nazywane są cechami **porządkowymi**. Następującym po sobie wariantom takich cech można przyporządkować kolejne liczby (rangi), odpowiadające zmieniającemu się natężeniu cechy. Cechy jakościowe, których warianty są określone w sposób opisowy i nie jest możliwe ustalenie dla nich hierarchii ważności (np. zawody: ekonomiczne, techniczne, humanistyczne) nazywane są cechami **nominalnymi**.

Pojęcie i metody badań statystycznych

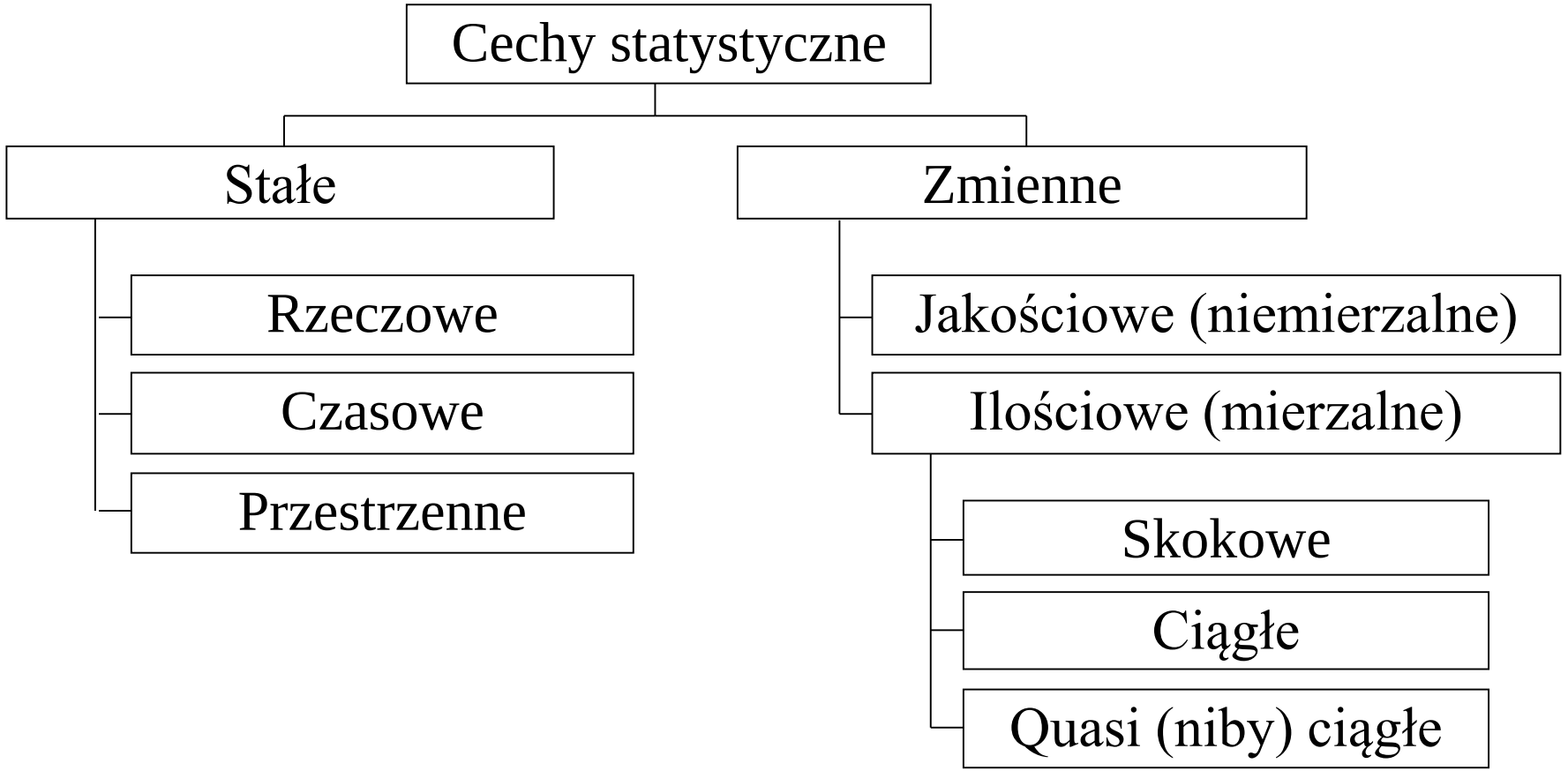
Wśród cech ilościowych (mierzalnych) wyróżnia się cechy **skokowe (dyskretne)** i **ciągłe**. Pierwsze z nich mogą przyjmować określone wartości ze skończonych i przeliczalnych przedziałów liczbowych, ale bez wartości pośrednich (np. liczba osób w gospodarstwie domowym, liczba pracowników firmy itp.). Warianty liczbowe tego rodzaju cech przyjmują zazwyczaj wartości liczb całkowitych. Cechy ilościowe **ciągłe** mogą natomiast przyjmować każdą wartość z określonego skończonego przedziału liczbowego (np. wiek, wzrost, waga). Cechy ciągłe są wyrażane za pomocą liczb rzeczywistych.

Trzecim rodzajem cechy ilościowej jest cecha **quasi (niby) ciągła**. W rzeczywistości jest to cecha skokowa, ale ze względu na wielką liczbę wariantów, jaką może przyjmować, w badaniach statystycznych traktuje się ją jako cechę ciągłą. Przykładami takich cech są: płace pracowników, ceny towarów, miesięczny dochód gospodarstw domowych, zysk przedsiębiorstwa itp. W obliczeniach statystycznych cechy mierzalne quasi – ciągłe zaliczane są do ciągłych.

Pojęcie i metody badań statystycznych

Podstawą podziału cech statystycznych na mierzalne i niemierzalne jest sposób ich pomiaru. Pomiarem nazywamy przyporządkowanie określonych symboli (liczb, nazw, znaków) własnościom mierzonych obiektów. Niejednokrotnie tę samą własność można zakwalifikować – w zależności od sposobu jej pomiaru – do cech mierzalnych lub niemierzalnych. Np. wiek osób mierzony w latach jest cechą mierzalną. Jeśli natomiast wiek badanych osób będziemy określać jako młody, średni lub podeszły – to będziemy mieli do czynienia z cechą niemierzalną.

Podział cech statystycznych



Źródło: B. Pułaska – Turyna: *Statystyka dla ekonomistów*, Difin, Warszawa 2005, s. 13

Pojęcie i metody badań statystycznych

Typ skali pomiarowej cech może być podstawą ich klasyfikacji na cztery grupy: **nominalne**, **porządkowe (rangowe)**, **przedziałowe (interwałowe)**, oraz **stosunkowe (ilorazowe)**. Stosownie do wymienionych wyżej rodzajów cech, wyróżnia się cztery skale **pomiarowe**: nominalną, porządkową, przedziałową i ilorazową.

Skala **nominalna** jest skalą najniższego poziomu. Liczby pełnią tutaj jedynie rolę umownych symboli służących do identyfikacji jednostek statystycznych i ich zaklasyfikowania do wyróżnionych kategorii. Na wartościach cech mierzonych w skali nominalnej można wykonywać tylko niektóre operacje matematyczne, takie jak: zliczanie, obliczanie częstości występowania, ustalenie klasy najliczniejszej. Przykładami liczb typu nominalnego są: numery na koszulkach zawodników drużyny piłki nożnej, numery tras autobusów, pokoi hotelowych, grupy krwi itp.

Pojęcie i metody badań statystycznych

Skala porządkowa (rangowa) jest wyższym poziomem pomiaru niż skala nominalna. Oprócz klasyfikacji umożliwia ona porządkowanie kategorii ze względu na natężenie danej cechy (np. wykształcenie: podstawowe, średnie, licencjackie, wyższe). Skala ta może ustalać porządek słaby lub mocny. W słabym uporządkowaniu dopuszczalne jest stwierdzenie „mniejszy lub równy" lub „większy lub równy". Nie można natomiast określić różnic między elementami ani stwierdzić, ile razy jeden element jest większy lub mniejszy od innego. W przypadku porządku mocnego: „większy niż" bądź „mniejszy niż" pojawia się asymetria. Przykładowo, jeśli X jest wyższy od Y, to nie może być na odwrót. Na liczbach w skali porządkowej można wykonywać takie same operacje jak dla skali nominalnej oraz dodatkowo – wyznaczać charakterystyki porządkowe i niektóre miary korelacji. Skala porządkowa posiada wszystkie własności skali nominalnej oraz dodatkowo umożliwia porządkowanie. Od skali porządkowej można przejść do nominalnej, ale nie odwrotnie. W skali porządkowej określa się np. gatunki owoców i warzyw (I, II, III), stopnie wojskowe, preferencje konsumentów.

Pojęcie i metody badań statystycznych

Skala przedziałowa (interwałowa) nie posiada prawdziwego punktu zerowego, tzw. „zera absolutnego”. Punkt ten jest ustalany arbitralnie (umownie). Przykładowo, w skali temperatury Celsjusza zerem jest punkt, w którym zamarza woda. Niższe temperatury zapisywane są ze znakiem minus, a wyższe – ze znakiem plus. W skali przedziałowej wyrażane są np. lata kalendarzowe, indeksy cen, skala temperatur Fahrenheita. Z uwagi na brak absolutnego punktu zerowego na liczbach w skali przedziałowej nie można wykonywać dzielenia. Wszystkie operacje dopuszczalne na skalach nominalnej i porządkowej są dozwolone na skali przedziałowej, a ponadto można tu obliczać średnie i miary zróżnicowania. Skala przedziałowa pozwala na określenie różnicy między poszczególnymi jednostkami. Oznacza to, że można tu nie tylko stwierdzić, że jedna wartość cechy jest mniejsza lub większa od drugiej, ale również określić czy różnica (dystans) między nimi jest taka sama (większa lub mniejsza) od różnicy między dwiema innymi wartościami. Niemożliwe jest natomiast stwierdzenie, ile razy jedna wartość jest większa lub mniejsza od drugiej.

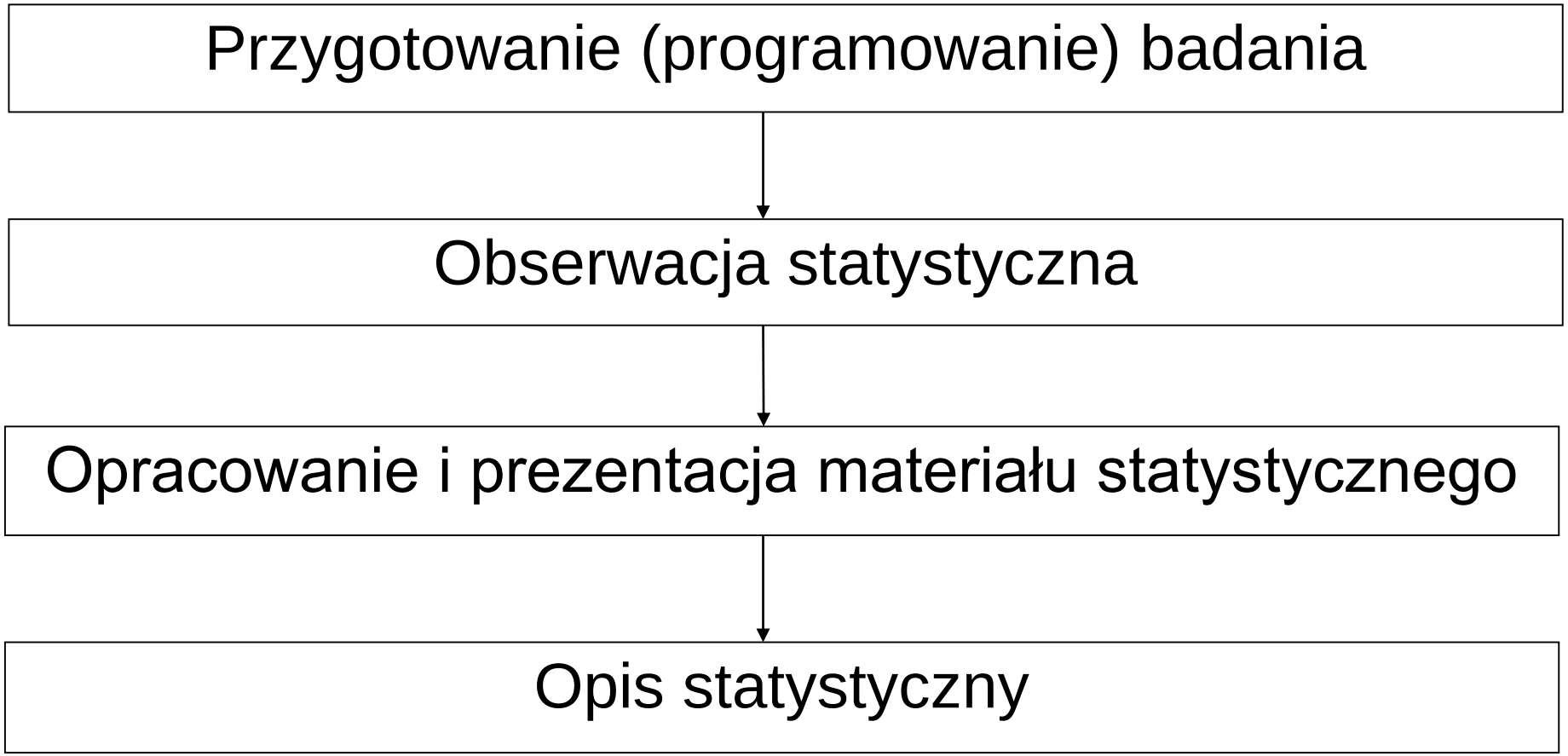
Pojęcie i metody badań statystycznych

Skala stosunkowa (ilorazowa) ma własności trzech poprzednich skal, a ponadto posiada naturalny punkt zerowy. Punkt ten oznacza brak danej cechy (fizyczny poziom zerowy). Umożliwia to dokonywanie na liczbach w tej skali wszystkich – łącznie z dzieleniem – operacji matematycznych. W skali stosunkowej mierzy się odległość, dochody, wielkość sprzedaży, ceny towarów, koszty itp.

Pojęcie i metody badań statystycznych

W wielowymiarowej analizie statystycznej użyteczną rolę pełni podział cech na stymulanty, destymulanty oraz nominanty . Zmienna jest stymulantą, jeśli wyższe jej wartości kwalifikują daną jednostkę statystyczną jako lepszą z punktu widzenia prowadzonego badania (np. dochody ludności w badaniach poziomu życia). W przypadku, gdy niższe wartości zmiennej stanowią o lepszej pozycji jednostki z punktu widzenia prowadzonych badań – to cecha ta jest destymulantą. Nominanta – to taka cecha, od której każde odchylenie (zarówno in plus, jak i in minus) jest zjawiskiem negatywnym z punktu widzenia prowadzonego badania (np. poziom substancji czynnej leku w organizmie człowieka w przypadku badań farmakologicznych). Wyróżnienie w zbiorze danych stymulant, destymulant i nominant stanowi punkt wyjścia przy liniowym porządkowaniu obiektów (np. województw, firm), czy wyodrębnianiu jednorodnych rejonów.

Etapy badania statystycznego



Etapy badania statystycznego

Badanie statystyczne to „zespół czynności zmierzających do uzyskania za pomocą metod statystycznych informacji charakteryzujących badana zbiorowość”. Przedmiotem statystyki opisowej są badania pełne (wyczerpujące, całkowite), w których biorą udział wszystkie jednostki danej zbiorowości statystycznej. Badaniami pełnymi są: **spis statystyczny**, **inwentaryzacja**, **rejestracja bieżąca** oraz **sprawozdawczość statystyczna**. **Spisem statystycznym** nazywamy okresowe lub doraźne specjalne badanie, obejmujące wszystkie jednostki danej zbiorowości w ściśle określonym czasie i przestrzeni. Jest on przeprowadzany za pomocą specjalnie przygotowanych formularzy. Formularz statystyczny jest odpowiednio przygotowanym zbiorem pytań i rubryk dotyczących badanych cech danej zbiorowości.

Etapy badania statystycznego

Do najważniejszych **spisów okresowych** przeprowadzanych w Polsce należy zaliczyć powszechne spisy ludności, spisy rolne, spisy pogłowia zwierząt gospodarskich itp. **Spisy doraźne** są przeprowadzane – w zależności od potrzeb – w różnych okresach (np. spisy środków trwałych). **Inwentaryzacja** jest spisem ujmującym faktyczny stan ilościowy i wartościowy majątku określonej jednostki gospodarczej lub administracyjnej. **Rejestracja bieżąca** polega na systematycznym notowaniu różnych faktów w momencie ich zaistnienia (np. rejestracja urodzeń, zgonów, małżeństw). **Sprawozdawczość statystyczna** obejmuje informacje statystyczne przekazywane obligatoryjnie na jednolitych formularzach przez osoby fizyczne, podmioty gospodarcze i instytucje do władz nadrzędnych lub organów statystycznych (np. zeznania podatkowe).

Pojęcie i metody badań statystycznych

W przypadku braku możliwości przeprowadzania badania całkowitego (jak również częściowego), stosuje się postępowanie zwane **szacunkiem statystycznym**. Polega ono na ustaleniu nieznanych wielkości lub właściwości zbiorowości na podstawie innych zbiorowości pozostających w określonym związku przyczynowo – skutkowym z interesującą nas nieznaną populacją.

Szczególną formą szacunku statystycznego jest **rachunek ekstrapolacyjny** lub **interpolacyjny**. **Ekstrapolacja** pozwala na ustalenie nieznanych wartości znajdujących się poza przedziałem wartości znanych. **Interpolacja** polega na szacowaniu nieznanych wartości badanej cechy na podstawie jej wcześniejszych lub późniejszych realizacji.

Pojęcie i metody badań statystycznych

Przygotowanie (programowanie) badania

Etap ten obejmuje czynności przygotowawcze, które zawierają: **ścisłe ustalenie celu i metody badania**, **określenie zbiorowości statystycznej i cech podlegających badaniu**, **zdefiniowanie jednostki statystycznej i jednostki sprawozdawczej**. Każde badanie statystyczne powinno być podporządkowane konkretnemu celowi, który leży poza statystyką. Po ustaleniu celu badania należy określić jego **przedmiot**, czyli **zbiorowość statystyczną** i **jednostkę badania**. Powinny być one ściśle zdefiniowane pod względem rzeczowym, czasowym i przestrzennym (terytorialnym).

Obserwacja statystyczna

Obserwacja statystyczna – w najogólniejszym ujęciu – polega na ustaleniu wartości cech ilościowych lub wariantów cech jakościowych u wszystkich jednostek tworzących badana zbiorowość. Może to odbywać się drogą uzyskania informacji od jednostek sprawozdawczych lub za pomocą bezpośredniego pomiaru. Zbiór danych uzyskanych w wyniku obserwacji tworzy **materiał statystyczny**.

Ze względu na źródło pochodzenia materiał statystyczny dzielimy na **pierwotny** i **wtórny**. Pierwszy z nich tworzą dane gromadzone specjalnie dla celów badania (np. dane zebrane w czasie spisu powszechnego czy dane pochodzące z Badań Aktywności Ekonomicznej Ludności – BAEL). Oznacza to, że badający uczestniczy w procesie gromadzenia danych pierwotnych. Dane pierwotne uzyskuje badacz przez obserwację bezpośrednią, wywiad, ankietę, eksperyment rynkowy itp.

Materiał **wtórny** tworzą dane, które zostały zgromadzone z myślą o innym problemie (na potrzeby innego badacza).

Pojęcie i metody badań statystycznych

Zebrane materiały statystyczne w swej pierwotnej postaci tworzą tzw. surowy materiał statystyczny, który jest – na ogół – obciążony błędami. Błędy te można podzielić na systematyczne (tendencyjne) oraz przypadkowe (niesystematyczne). Błędy systematyczne wynikają z jednokierunkowej tendencji do zniekształcania badanej rzeczywistości. W dużej liczbie spostrzeżeń tego rodzaju błędy sumują się, w rezultacie czego dane sumaryczne są wielkościami zawyżonymi lub zaniżonymi w stosunku do stanu faktycznego. Błędy przypadkowe są popełniane nieumyślnie i mogą wynikać z nieuwagi, z nieumiejętności podania prawidłowych informacji bądź też ze zwykłego niedbalstwa.

Pojęcie i metody badań statystycznych

W celu wykrycia ewentualnych błędów i nieścisłości surowy materiał statystyczny poddaje się kontroli formalnej i merytorycznej. Zadaniem kontroli formalnej (ilościowej) jest sprawdzenie kompletności, pełności i zupełności danego materiału statystycznego. Kontrola merytoryczna obejmuje kontrolę logiczną i arytmetyczną. Pierwsza z nich polega na sprawdzeniu, czy treść rubryk formularza statystycznego odpowiada rzeczywistości (np. może się zdarzyć, że w formularzu podano poziom płac brutto zamiast netto, a są to odmienne pojęcia i różne liczby). Celem kontroli arytmetycznej jest natomiast porównanie liczb otrzymanych z sumowania danych źródłowych z liczbami w sprawozdaniach zbiorczych.

Opracowanie i prezentacja materiału statystycznego

Podstawowym problemem z dużymi zbiorami danych jest nadmiar informacji. Z tego też względu niezbędną czynnością jest opracowanie i odpowiednia prezentacja materiału statystycznego. Pozwala to z jednej strony na uporządkowanie zebranych danych, z drugiej zaś na znaczącą – ale nie nadmierną – redukcję informacji.

Jedną z czynności związanych z usystematyzowaniem materiału statystycznego jest grupowanie. Grupowanie polega na podziale zbiorowości na możliwie jednorodne grupy zgodnie z przyjętymi kryteriami.

Opracowanie i prezentacja materiału statystycznego

Zgodnie z zasadami logiki formalnej, poprawnie przeprowadzone grupowanie powinno być:

- **wyczerpujące**, tj. każda jednostka zbiorowości musi być sklasyfikowana i włączona do odpowiedniej podgrupy,
- **rozłączne**, tzn. wyróżnione podgrupy muszą wzajemnie wykluczać się,
- **efektywne**, czyli wyróżnione podgrupy winny być na tyle jednorodne jakościowo, by stanowiły podstawę do formułowania uogólniających twierdzeń.

W wyniku grupowania statystycznego według jednej cechy powstają **szeregi statystyczne**, jeśli natomiast grupowania dokonujemy według dwóch lub większej ilości cech, tworzone są **tablice statystyczne**.

ORGANIZACJA BADAŃ STATYSTYCZNYCH

PROCES BADANIA STATYSTYCZNEGO

Etapy:

I. przygotowanie,

II. obserwacja statystyczna,

materiał statystyczny: ***pierwotny, wtórny,***

metody obserwacji statystycznej: ***pełna*** **(spis, rejestracja bieżąca, sprawozdawczość),**

częściowa **(reprezentacyjna, monografia, ankieta).**

Metody obserwacji statystycznej mogą mieć charakter *ciągły* (np. rejestracja urodzeń, zgonów, pojazdów, ...), *okresowy* (np. spis powszechny), doraźne (np., sondaż na okoliczność, spis na okoliczność).

III. opracowanie

- **kontrola:** ***formalna – polega na sprawdzeniu kompletności uzyskanych formularzy oraz zapisów***

merytoryczna – polega na eliminowaniu (lub przynajmniej minimalizowaniu) błędów przypadkowych i systematycznych.

III. Opracowanie c.d.

!!! Najbardziej niebezpieczne dla poprawności wniosków są błędy systematyczne (jednokierunkowe), na przykład zaniżanie dochodów czy zawyżanie wydatków na kulturę. Należy zdawać sobie sprawę z rzędu wielkości takiego błędu i jeśli przewidujemy, że może być on wysoki należy odpowiedzi szczegółowe (wartości) zastąpić przedziałami.

ORGANIZACJA BADAŃ STATYSTYCZNYCH

III. Opracowanie c.d.

- grupowanie,

- ***typologiczne (jakościowe)*** **– polega na wydzieleniu jednorodnych grup z niejednorodnej zbiorowości na podstawie wariantów badanej cech lub cech,**
- ***wariancyjne*** **– oparte na cesze ilościowej – warianty danej cechy nie różnią się jakościowo w istotny sposób,**

!!! Grupując rezygnujemy ze szczegółów, ale podkreślamy wspólne właściwości jednostek badanej zbiorowości.

- prezentacja,

- ✓ ***szeregi statystyczne,***
- ✓ ***tablice statystyczne,***
- ✓ ***wykresy statystyczne.***

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

ROZKŁAD EMPIRYCZNY **to rozkład cechy, w którym poszczególnym wariantom (wartościom) cechy podanym punktowo lub przedziałowo przyporządkowane są liczebności z jakimi dane warianty cechy w badanej zbiorowości występują.**

ROZKŁAD EMPIRYCZNY **otrzymuje się w wyniku badania statystycznego i najczęściej prezentuje się w formie szeregu statystycznego, tablicy statystycznej lub wykresu statystycznego.**

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

Do oceny rozkładu empirycznego badanej zbiorowości wykorzystuje się szereg różnych charakterystyk liczbowych (mierników statystycznych), które mogą mieć charakter liczb absolutnych albo względnych.

Liczby absolutne (bezwzględne) są to wielkości, które otrzymuje się w wyniku mierzenia lub sumowania jednostek badanej zbiorowości. Mają one charakter wielkości pierwotnych, są liczbami mianowanymi i wyraża się je w sztukach, litrach, metrach, kilogramach, hektarach, złotówkach itp.., itd.

Liczby względne są to wielkości uzyskane w drodze dzielenia przez siebie dwóch liczb absolutnych. Przedstawiają one stosunki pomiędzy liczbami absolutnymi.

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

Prezentacji danych statystycznych można dokonać przy pomocy:

a) szeregu statystycznego,

b) tablicy statystycznej,

c) wykresu statystycznego.

Ad. a)

KONSTRUKCJA SZEREGÓW STATYSTYCZNYCH

Oznaczenia:

x_i – i-ta wartość (wariant) cechy, gdzie i = 1, ..., n lub i = 1, ..., k

n_i – liczebność i-tego wariantu cechy,

n - liczebność badanej zbiorowości statystycznej,

k - liczba przedziałów (klas) – dotyczy szeregów rozdzielczych.

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

Szeregi statystyczne

Szeregiem statystycznym nazywamy ciąg wielkości statystycznych (danych statystycznych, materiał statystyczny) uporządkowany według określonego kryterium, tzn. rosnąco lub malejąco:

według kryterium formalnego wyróżnia się szeregi:

- **szczegółowe (wyliczające),**
- **rozdzielcze punktowe,**
- **rozdzielcze przedziałowe,**
- **skumulowane,**

według kryterium merytorycznego wyróżnia się szeregi:

- **strukturalne,**
- **czasowe,**
- **przestrzenne.**

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

Rodzaje szeregów statystycznych

1. **Szereg szczegółowy (wyliczający)**

Szereg szczegółowy to uporządkowany rosnąco lub malejąco, ale nie pogrupowany materiał statystyczny.

$x_i : x_1, x_2, \ldots, x_n \qquad i = 1, \ldots, n$

gdzie: x_i - warianty (wartości) cechy

2. Szereg rozdzielczy punktowy

Szereg rozdzielczy punktowy to pogrupowany materiał statystyczny według wartości cechy, którym odpowiada liczba jednostek (liczebność) posiadających dany wariant cechy.

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

Szereg rozdzielczy punktowy

x_i	n_i
x_1	n_1
x_2	n_2
.....	
x_k	n_k
	n

$i = 1, \ldots, k$

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

3. Szereg rozdzielczy przedziałowy

Szereg rozdzielczy przedziałowy to pogrupowany materiał statystyczny według wariantów cechy podanych przedziałowo, którym odpowiada liczba jednostek (liczebność) posiadających wariant cechy pochodzący z danego przedziału.

x_{0i}-x_{1i}	x_{01}-x_{11}	x_{02}-x_{12}	...	x_{0k}-x_{1k}
n_i	n_1	n_2	...	n_k

Szereg rozdzielczy przedziałowy

x_{0i} - x_{1i}	n_i
x_{01} - x_{11}	n_1
x_{02} - x_{12}	n_2
.....	
x_{0k} - x_{1k}	n_k
	n

i = 1, . . . , k

Elementy szeregu rozdzielczego przedziałowego:

x_{0i} - dolna granica i-tego przedziału

x_{1i} - górna granica i-tego przedziału

h_i - rozstęp (rozpiętość, interwał) i-tego przedziału wyznaczany według wzoru: $h_i = x_{1i}-x_{0i}$

$\hat{x}_i$ - środek i-tego przedziału wyznaczany według wzoru $(x_{0i}+x_{1i}):2$

W szeregach rozdzielczych do określenia rozkładu (struktury) obok liczebności bezwzględnej stosuje się także odpowiedni *wskaźnik struktury*.

WSKAŹNIK STRUKTURY (liczebność względna, częstość, frakcja, odsetek) to stosunek liczby jednostek o danej wartości cechy do liczebności próby:

$$\omega_i = \frac{n_i}{n}, \text{ przy czym } \sum_{i=1}^{k} \omega_i = 1, \quad 0 \le \omega_i \le 1.$$

U!!! Gdy liczebność próby jest niewielka (n < 30), przy interpretacji rozkładu nie powinno się w zasadzie korzystać z częstości, a jedynie liczebności.

Ustalenie liczby przedziałów (w przybliżeniu)

$k = n^{1/2}$, $k = 1 + 3,322 \log n$, $k = 5 \lg n$

Ustalenie rozpiętości przedziałów:

$h_i = (x_{max} – x_{min}) : k = R : k$

$R = x_{max} – x_{min}$ – rozstęp , k – liczba klas

UWAGA! $hk \geq R$

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

Przy grupowaniu z reguły staramy się ustalić jednakowe rozpiętości przedziałów klasowych. Przy równej rozpiętości przedziałów klasowych liczebności (częstości) występujące w poszczególnych klasach są porównywalne. Ale jeśli badana populacja jest niejednorodna i występuje duża koncentracja wartości w jednej grupie, wówczas należy stosować różne rozpiętości przedziałów klasowych.

W celu oceny struktury badanej zbiorowości przy różnych rozpiętościach przedziałów klasowych konieczne jest stosowanie zamiast liczebności wskaźnika zwanego gęstością liczebności.

Gęstość liczebności (częstości) jest to stosunek liczebności (częstości) danej klasy do rozpiętości przedziału klasowego:

gęstość liczebności: $f_l = \frac{n_i}{h_i},$ **gęstość częstości:** $f_\omega = \frac{\omega_i}{h_i},\ \mathrm{i} = 1, 2, ..., \mathrm{k}$

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

4. Szereg skumulowany:

- **rosnąco**

x_{0i}-x_{1i}	x_{01}-x_{11}	x_{02}-x_{12}	...	x_{0k}-x_{1k}
$n\uparrow$	n_1	n_1+n_2	...	$n_1+n_2+ \dots +n_k$

- **malejąco**

x_i	x_1	x_2	...	x_k
$n\downarrow$	$n_k+n_{k-1}+ \dots +n_1$	$n_k+n_{k-1}+ \dots +n_2$	...	n_k

---

L.p.	Wartość
1	3,1
2	2,4
3	3,5
4	2,5
5	2,6
6	2,2
7	3,3
8	2,1
9	4,1
10	4,0

L.p.	Wartość
1	2,1
2	2,2
3	2,4
4	2,5
5	2,6
6	3,1
7	3,3
8	3,5
9	4,0
10	4,1

L.p.	Wartość
1	2,1
2	4,3
3	6,7
4	9,2
5	11,8
6	14,9
7	18,2
8	21,7
9	25,7
10	29,8

L.p.	Wartość
1	4,1
2	4,0
3	3,5
4	3,3
5	3,1
6	2,6
7	2,5
8	2,4
9	2,2
10	2,1

L.p.	Wartość
1	4,1
2	8,1
3	11,6
4	14,9
5	18,0
6	20,6
7	23,1
8	25,5
9	27,7
10	29,8

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

Szereg przestrzenny (terytorialny, geograficzny)

Szeregiem przestrzennym nazywa się taki szereg statystyczny, który obrazuje badane zjawisko na danym obszarze (terytorium).

Szereg czasowy (dynamiczny, rozwojowy)

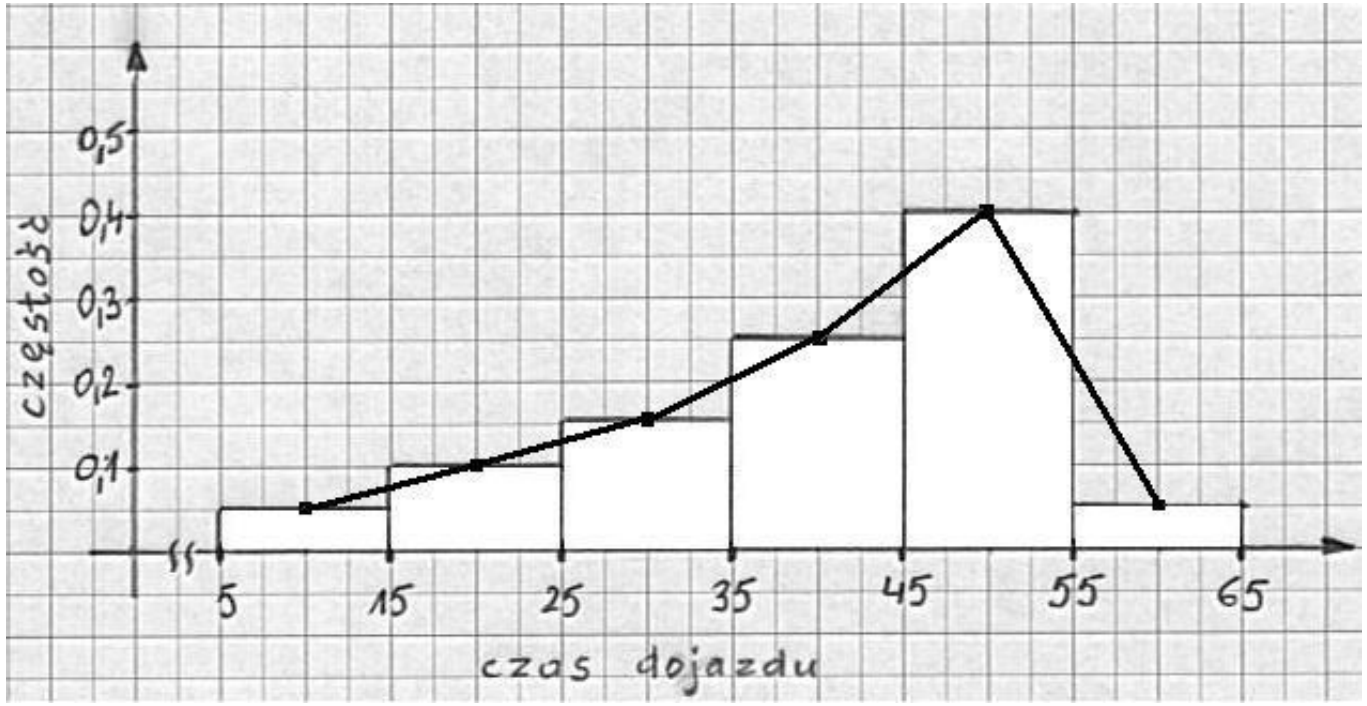
Szeregiem czasowym nazywa się taki szereg statystyczny, który przedstawia czasowe zmiany badanego zjawiska.

Rozróżniamy dwa rodzaje szeregów czasowych:

1. **Szeregi czasowe okresów** – zawierają informacje o rozmiarach zjawiska w określonych okresach (miesiąc, kwartał, rok) np. produkcja w danej branży w kwartałach,
2. **Szeregi czasowe momentów** - ujmuje wielkość zjawiska w danym momencie czasu, np. stan zapasów magazynowych danego materiału na 31.03.2006 r.

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

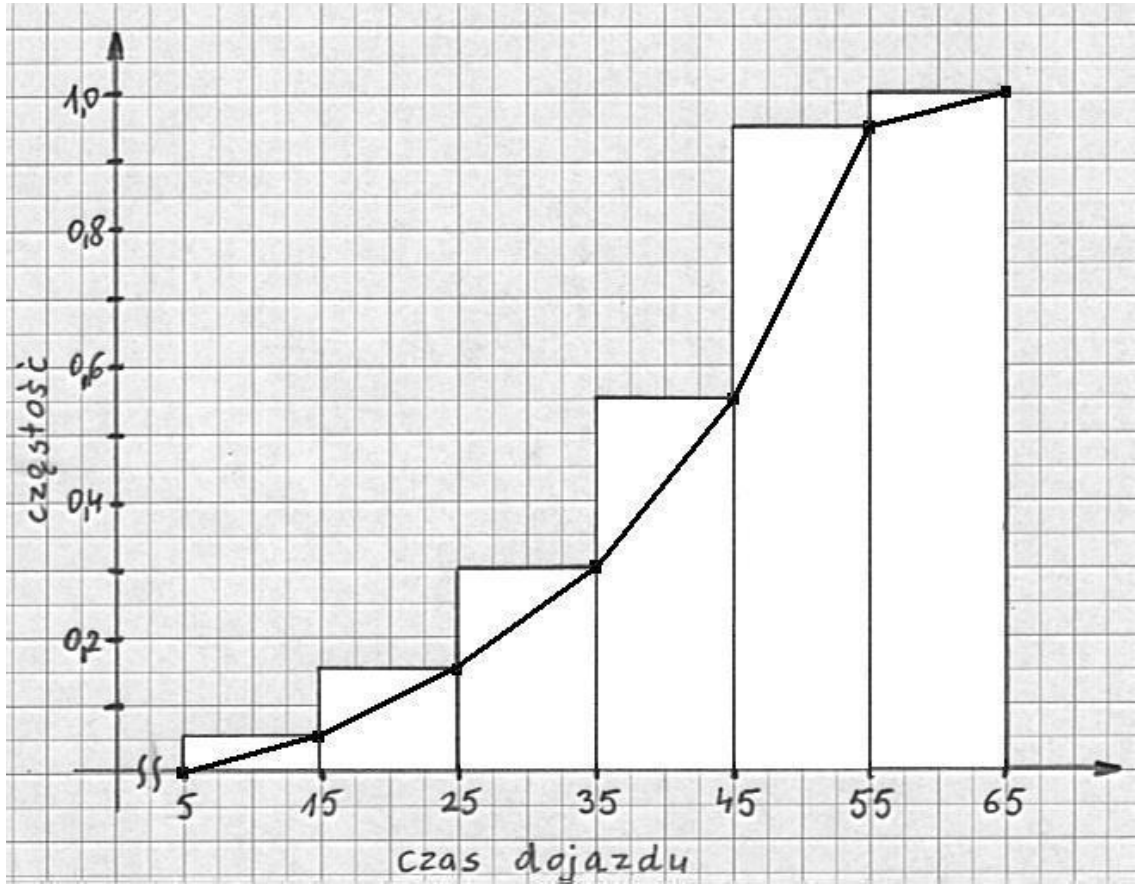
HISTOGRAM **DIAGRAM**



DYSTRYBUANTĄ EMPIRYCZNĄ nazywamy przyporządkowanie kolejnym wartościom cechy statystycznej (zmiennej) odpowiadających im częstości skumulowanych (ewentualnie liczebności skumulowanych).

OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

Histogram i diagram częstości skumulowanej



OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

Ad. b)**TABLICA STATYSTYCZNA (MAKIETA)**

Tablica nr . Tytuł tablicy

Tytuł boczka	Tytuł główki			Razem (Ogółem)
Boczek tablicy		. . . . .		
Razem (Ogółem)				

Główka tablicy

Źródło: ……

Tablice statystyczne możemy podzielić na:
-proste,
- złożone,
- kombinowane

- robocze,
- wynikowe,

- szczegółowe,
- ogólne,

-oparte na:
✓ liczbach absolutnych,
✓ liczbach względnych.

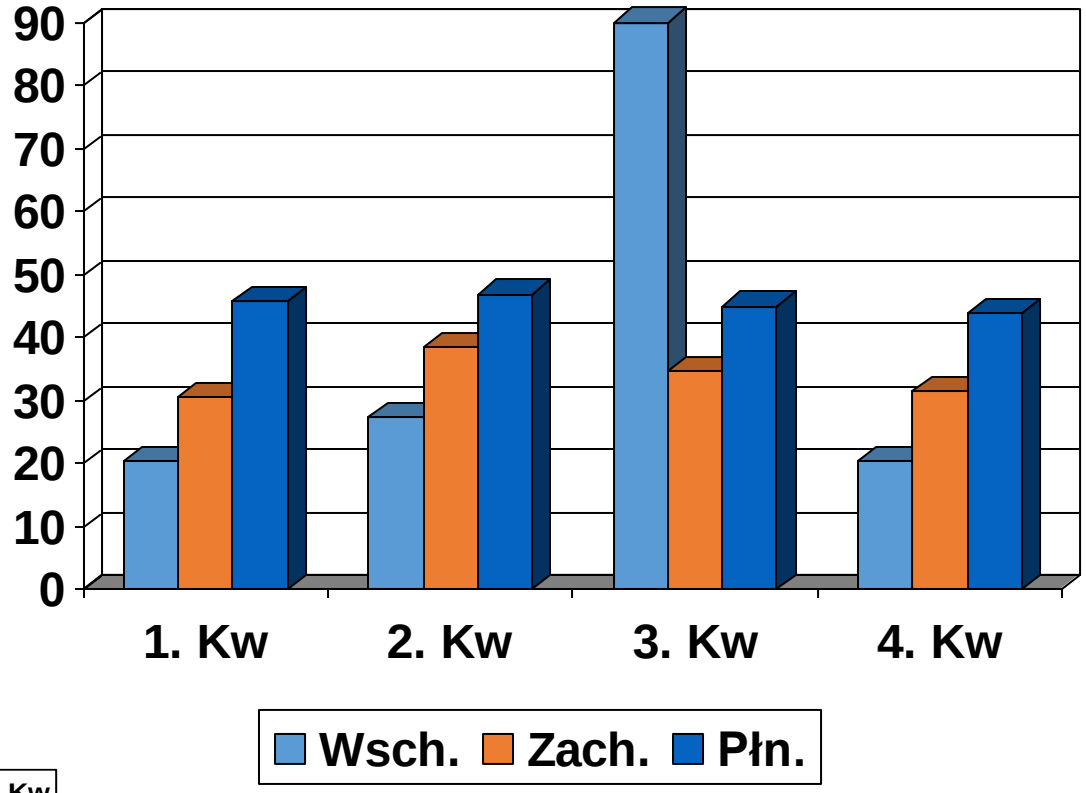
Znaki umowne
(-) - zjawisko nie występuje
(.) - brak danych, lub brak danych wiarygodnych
(0) - zjawisko występuje, ale w ilościach mniejszych niż przyjęto w tablicy
(x) - wypełnienie pozycji nie jest możliwe lub celowe ze względu na układ tabeli

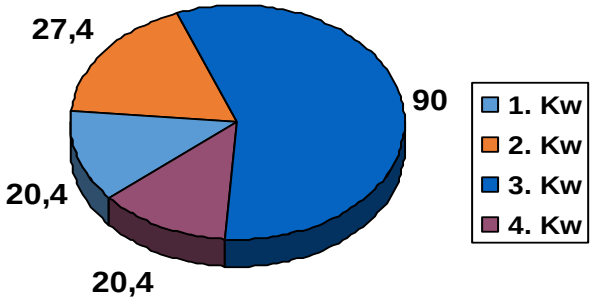
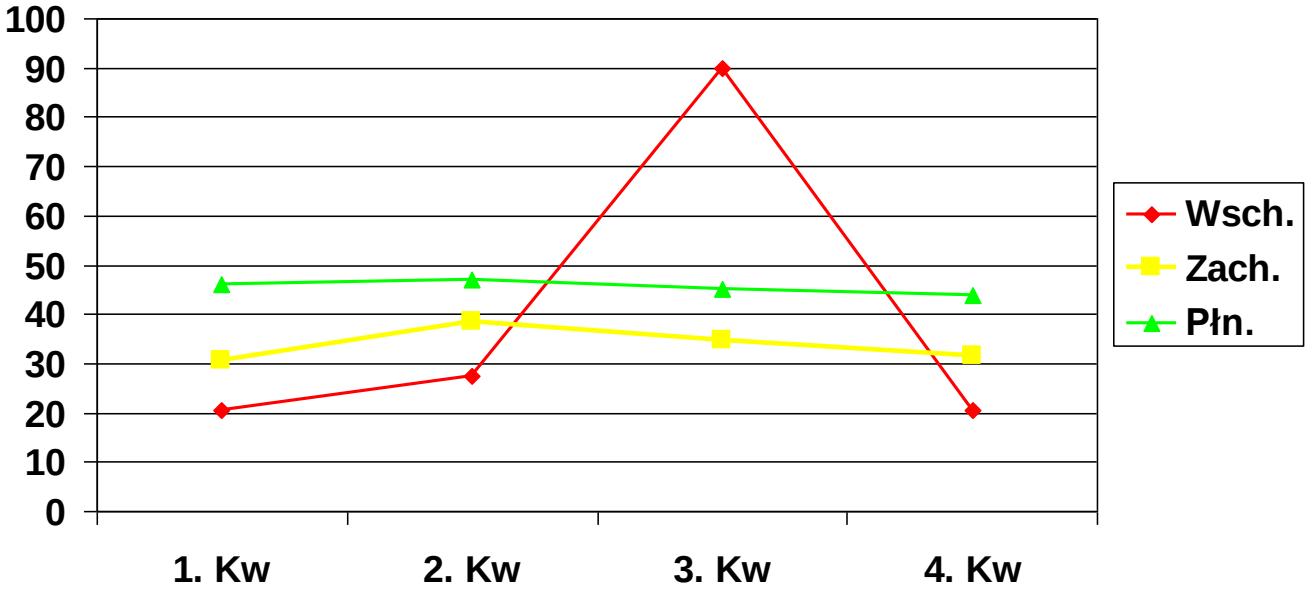
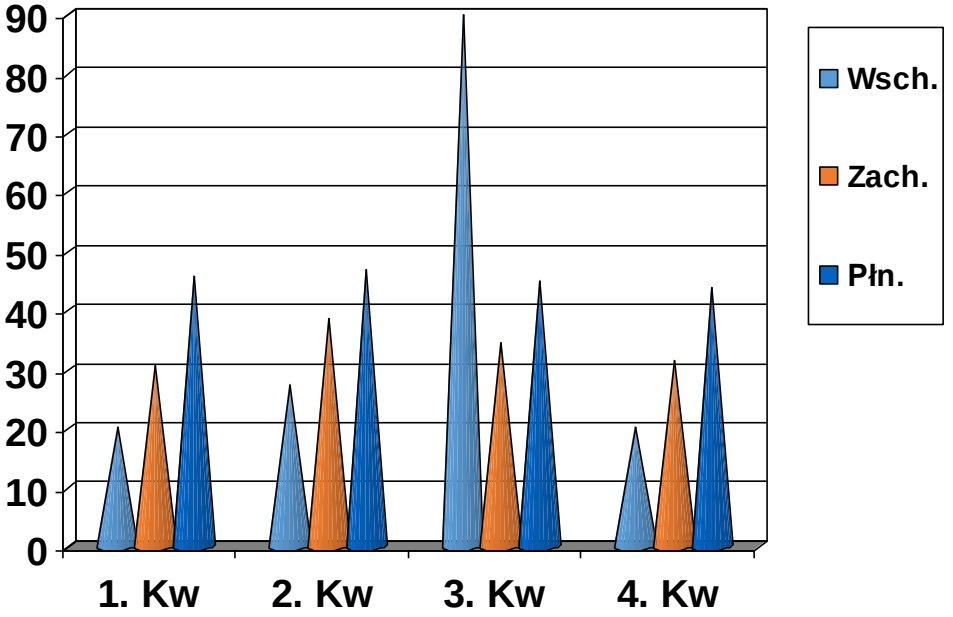
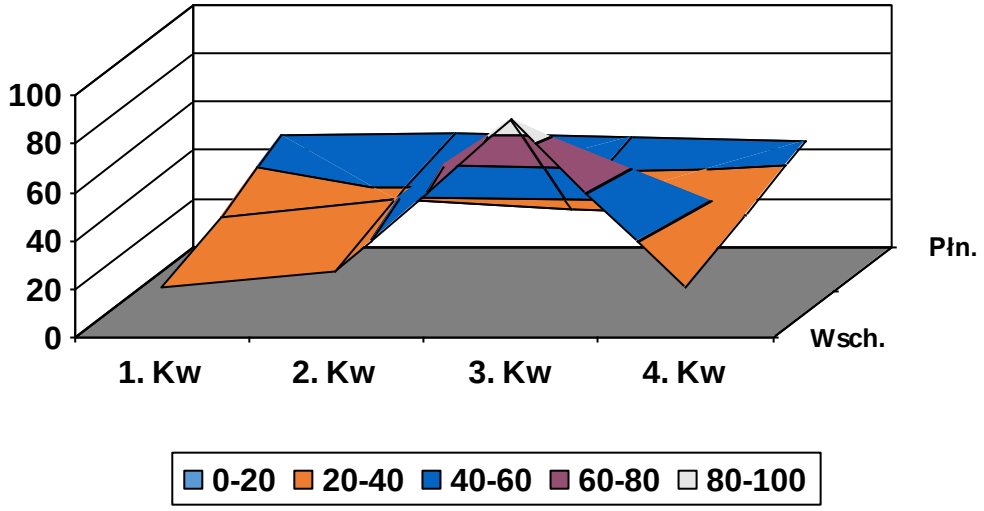
OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO

WYKRES JEST WIZUALNĄ FORMĄ REJESTRACJI SZCZEGÓŁOWYCH DANYCH oraz NARZĘDZIEM PREZENTACJI I ANALIZY UOGÓLNIONYCH INFORMACJI STATYSTYCZNYCH.

WYKRESY STATYSTYCZNE

- **liniowe,**
- **kołowe, słupkowe,**
- **histogramy,**
- **diagramy,**
- **obrazkowe,**
- **piramidy,**
- **krzywa liczebności,**
- **kartogramy,**
- **… i wiele innych.**





OPRACOWANIE I PREZENTACJA MATERIAŁU STATYSTYCZNEGO